(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 112602077 A
(43)申请公布日 2021.04.02

(21)申请号 201980035900.0

(22)申请日 2019.04.03

(30)优先权数据
15/991,438 2018.05.29 US

(85)PCT国际申请进入国家阶段日
2020.11.27

(86)PCT国际申请的申请数据
PCT/US2019/025638 2019.04.03

(87)PCT国际申请的公布数据
WO2019/231559 EN 2019.12.05

(71)申请人 索尼互动娱乐有限责任公司
地址 美国加利福尼亚州

(72)发明人 F.罗贾斯-埃切尼奎 M.斯乔林 U.默特 S.谢克 M.K.奇特拉

(74)专利代理机构 北京市柳沈律师事务所 11105
代理人 张晓明

(51)Int.Cl.
G06F 16/78(2019.01)
G06F 16/75(2019.01)
H04N 21/466(2011.01)
H04N 21/472(2011.01)

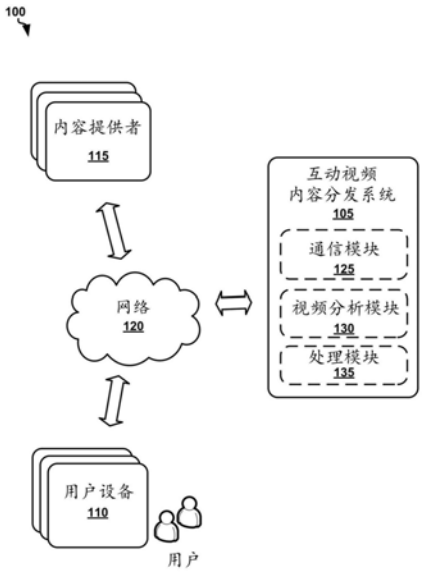
权利要求书4页 说明书11页 附图6页

(54)发明名称

交互式视频内容分发

(57)摘要

本发明提供用于交互式视频内容分发的方法和系统。一种示例性方法包括:接收诸如直播电视或视频流的视频内容。该方法可以对视频内容的视频帧运行一个或多个机器学习分类器,以创建与机器学习分类器相对应的分类元数据、以及与分类元数据相关联的一个或多个概率得分。此外,该方法可以基于一组预定规则和可选的用户简档来创建一个或多个交互触发器。该方法可以确定用于触发至少一个触发器的条件被满足,并基于所述确定、分类元数据和概率得分来触发关于视频内容的至少一个动作。例如,该动作可以分发附加信息、呈现建议、自动编辑视频内容或控制视频内容的分发。



CN 112602077 A

1.一种用于交互式视频内容分发的系统,所述系统包括:

通信模块,其被配置为接收视频内容,所述视频内容包括一个或多个视频帧;

视频分析器模块,其被配置为对所述一个或多个视频帧运行一个或多个机器学习分类器,以创建分类元数据以及与所述分类元数据相关联的一个或多个概率得分,所述分类元数据对应于所述一个或多个机器学习分类器;以及

处理模块,其被配置为基于规则集创建一个或多个交互触发器,所述一个或多个交互触发器被配置为基于所述分类元数据触发与所述视频内容有关的一个或多个动作。

2.一种用于交互式视频内容分发的方法,所述方法包括:

通过通信模块接收视频内容,所述视频内容包括一个或多个视频帧;

通过处理模块对所述一个或多个视频帧运行一个或多个机器学习分类器,以创建分类元数据以及与所述分类元数据相关联的一个或多个概率得分,所述分类元数据对应于所述一个或多个机器学习分类器;以及

通过处理模块基于规则集创建一个或多个交互触发器,所述一个或多个交互触发器被配置为基于所述分类元数据触发与所述视频内容有关的一个或多个动作。

3.根据权利要求1所述的方法,其中所述一个或多个动作的触发还基于所述一个或多个概率得分。

4.根据权利要求1所述的方法,其中所述视频内容包括实时视频,所述实时视频被延迟直到对所述一个或多个视频帧运行所述一个或多个机器学习分类器。

5.根据权利要求1所述的方法,其中所述视频内容包括点播视频,在所述视频内容被上载到内容分发网络(CDN)之前,对所述一个或多个视频帧运行所述一个或多个机器学习分类器。

6.根据权利要求1所述的方法,其中所述视频内容包括视频游戏。

7.根据权利要求1所述的方法,还包括:

确定用于触发所述一个或多个交互触发器中至少一个的条件被满足;以及

响应于所述确定,触发与所述视频内容有关的所述一个或多个动作。

8.根据权利要求1所述的方法,其中所述一个或多个机器学习分类器包括图像识别分类器,所述图像识别分类器被配置为分析一个所述视频帧中的静态图像,并且其中,所述一个或多个机器学习分类器包括复合识别分类器,所述复合识别分类器被配置为分析:(i)两个或多个所述视频帧之间的一个或多个图像变化;和(ii)两个或多个所述视频帧之间的一个或多个声音变化。

9.根据权利要求1所述的方法,还包括:创建与所述一个或多个交互触发器相对应的一个或多个入口点,其中所述一个或多个入口点中的每个包括与所述视频内容相关联的用户输入或与所述视频内容相关联的用户姿势。

10.根据权利要求9所述的方法,其中所述一个或多个入口点中的每个包括以下一个或多个:所述视频内容的暂停、所述视频内容的跳转点、所述视频内容的书签、所述视频内容的位置标记、与所述视频内容相关联的搜索结果、以及语音命令。

11.根据权利要求9所述的方法,其中所述一个或多个动作基于与所述视频内容的一个所述入口点相关联的帧的所述分类元数据。

12.根据权利要求1所述的方法,其中所述规则集基于以下一个或多个:用户简档、用户

设置、用户偏好、观看者身份、观看者年龄和环境条件。

13.根据权利要求1所述的方法,其中:

所述一个或多个机器学习分类器包括一般对象分类器,所述一般对象分类器被配置为识别存在于所述一个或多个视频帧中的一个或多个对象;并且

在触发所述一个或多个交互触发器时要采取的所述一个或多个动作包括以下一个或多个:用所述一个或多个视频帧中的新对象替换所述一个或多个对象、自动突出显示所述对象、推荐由所述一个或多个对象表示的可购买项目、基于对所述一个或多个对象的识别编辑所述视频内容、基于对所述一个或多个对象的识别控制所述视频内容的分发、以及呈现与所述一个或多个对象相关的搜索选项。

14.根据权利要求1所述的方法,其中:

所述一个或多个机器学习分类器包括产品分类器,所述产品分类器被配置为识别存在于所述一个或多个视频帧中的一个或多个可购买项目;以及

在触发所述一个或多个交互触发器时要采取的所述一个或多个动作包括:提供一个或多个链接,使得用户能够购买所述一个或多个可购买项目。

15.根据权利要求1所述的方法,其中:

所述一个或多个机器学习分类器包括环境条件分类器,所述环境条件分类器被配置为确定与所述一个或多个视频帧相关联的环境条件;

基于以下传感器数据创建所述分类元数据:一个或多个观看者正在观看所述视频内容的场所的照明条件、所述场所的噪声水平、与所述场所相关联的观众观看者类型、观看者身份、当前时间,其中使用一个或多个传感器获得所述传感器数据;并且

在触发所述一个或多个交互触发器时要采取的所述一个或多个动作包括以下一个或多个:基于所述环境条件编辑所述视频内容、基于所述环境条件控制所述视频内容的分发、基于所述环境条件提供与所述视频内容或另一媒体内容相关联的建议、以及提供与所述环境条件相关联的另一媒体内容。

16.根据权利要求1所述的方法,其中:

所述一个或多个机器学习分类器包括情绪条件分类器,所述情绪条件分类器被配置为确定与所述一个或多个视频帧相关联的情绪水平;

基于以下一个或多个来创建所述分类元数据:所述一个或多个视频帧的颜色信息、所述一个或多个视频帧的音频信息、用户在观看所述视频内容时表现出的用户行为;并且

在触发所述一个或多个交互触发器时要采取的所述一个或多个动作包括以下一个或多个:提供关于与所述情绪水平相关联的另一媒体内容的建议、以及提供与所述情绪水平相关联的另一媒体内容。

17.根据权利要求1所述的方法,其中:

所述一个或多个机器学习分类器包括地标分类器,所述地标分类器被配置为识别存在于在所述一个或多个视频帧中的地标;并且

在触发所述一个或多个交互触发器时要采取的所述一个或多个动作包括以下一个或多个:在所述一个或多个视频帧中标记所识别的地标、提供关于与所识别的地标相关联的另一媒体内容的建议、提供与所识别的地标相关联的另一媒体内容、基于所识别的地标编辑所述视频内容、基于所识别的地标控制所述视频内容的分发、以及呈现与所识别的地标

相关的搜索选项。

18.根据权利要求1所述的方法,其中:

所述一个或多个机器学习分类器包括人物分类器,所述人物分类器被配置为识别存在于所述一个或多个视频帧中的一个或多个个体;并且

在触发所述一个或多个交互触发器时要采取的所述一个或多个动作包括以下一个或多个:在所述一个或多个视频帧中标记所述一个或多个个体、提供关于与所述一个或多个个体相关联的另一媒体内容的建议、提供与所述一个或多个个体相关联的另一媒体内容、基于所述一个或多个个体编辑所述视频内容、基于所述一个或多个个体控制所述视频内容的分发、以及呈现与所述一个或多个个体相关的搜索选项。

19.根据权利要求1所述的方法,其中:

所述一个或多个机器学习分类器包括食物分类器,所述食物分类器被配置为识别存在于所述一个或多个视频帧中的一个或多个食物项;并且

在触发所述一个或多个交互触发器时要采取的所述一个或多个动作包括以下一个或多个:在所述一个或多个视频帧中标记所述一个或多个食物项、提供与所述一个或多个食物项相关的营养信息、为用户提供购买与所述一个或多个食物项相关联的可购买项目的购买选项、提供与所述一个或多个食物项关联的媒体内容、以及提供与所述一个或多个食物项相关的搜索选项。

20.根据权利要求1所述的方法,其中:

所述一个或多个机器学习分类器包括问题内容分类器,所述可以内容分类器被配置为检测所述一个或多个视频帧中的问题内容,所述问题内容包括以下一个或多个:裸体、武器、酒精、烟草、毒品、血液、仇恨言论、亵渎、血腥、以及暴力;并且

在触发所述一个或多个交互触发器时要采取的所述一个或多个动作包括以下一个或多个:在向用户显示之前自动模糊所述一个或多个视频帧中的所述问题内容、跳过与所述问题内容相关联的所述视频内容的部分、基于所述问题内容编辑所述视频内容、基于所述问题内容调整所述视频内容的音频、基于所述问题内容调整音频音量水平、基于所述问题内容控制所述视频内容的分发、以及将所述问题内容通知用户。

21.一种用于交互式视频内容分发的系统,所述系统包括:

通信模块,其接收视频内容,所述视频内容包括一个或多个视频帧;

视频分析器模块,其对所述一个或多个视频帧运行一个或多个机器学习分类器,以创建一个或多个分类元数据集以及与所述一个或多个分类元数据集相关联的一个或多个概率得分,所述一个或多个分类元数据集对应于所述一个或多个机器学习分类器;以及

处理模块,其基于规则集创建一个或多个交互触发器,所述一个或多个交互触发器被配置为基于所述一个或多个分类元数据集触发与所述视频内容有关的一个或多个动作。

22.一种非暂时性处理器可读介质,其上存储有指令,当由一个或多个处理器执行时,所述指令使所述一个或多个处理器实现用于跳过媒体内容的一个或多个不需要的部分的方法,所述方法包括:

通信模块,其被配置为接收视频内容,所述视频内容包括一个或多个视频帧;

视频分析器模块,其被配置为对所述一个或多个视频帧运行一个或多个机器学习分类器,以创建对应于所述一个或多个机器学习分类器的分类元数据、以及与所述分类元数据

相关联的一个或多个概率得分;以及

处理模块,其被配置为基于规则集创建一个或多个交互触发器,所述一个或多个交互触发器被配置为基于所述分类元数据触发与所述视频内容有关的一个或多个动作。

交互式视频内容分发

技术领域

[0001] 本公开一般地涉及视频内容处理，更具体地，涉及用于交互式视频内容分发的方法和系统，其中可以基于机器学习分类器创建的分类元数据触发各种动作。

背景技术

[0002] 本节所述的方法可以采用，但不一定是先前设想或采用的方法。因此，除非另有说明，否则不应假设本节所述的任何方法仅因其包含在本节中而被视为现有技术。

[0003] 电视节目、电影、通过点播视频获得的视频、计算机游戏、以及其他媒体内容可以通过互联网、空中广播、线缆、卫星或蜂窝网络来分发。电子媒体设备(如用户家中的电视显示器、个人计算机或游戏机)具有接收、处理和显示媒体内容的能力。现代用户面临着大量的媒体内容选项，这些内容随时都可以使用。然而，许多用户发现很难与媒体内容交互(例如，选择附加的媒体内容或了解通过媒体内容呈现的某些对象的更多信息)。

发明内容

[0004] 提供此发明内容是为了以简化形式介绍一系列概念，这些概念将在下面的具体实施方式中进一步描述。本发明内容并不旨在标识所要求保护的主题的关键特征或基本特征，也不意图用于帮助确定所要求保护的主题的范围。

[0005] 本公开涉及交互式视频内容分发。该技术用于：接收视频内容，诸如直播电视、视频流或用户生成的视频；分析视频内容的每一帧以确定相关的分类；以及基于分类触发动作。这些动作可以提供附加信息、呈现建议、编辑视频内容或控制视频内容分发等。提供多个机器学习分类器来分析每个缓冲帧，以动态和自动地创建表示视频内容中的一个或多个资产(asset)的分类元数据。一些示例性资产包括出现在视频内容中的个体或地标、各种预定对象、食物、可购买项目、视频内容类型、关于观看视频内容的观众的信息、环境条件等。用户可能会对被触发的动作做出反应，这可能会改善他们的娱乐体验。例如，用户可以搜索关于出现在视频内容中的演员的信息，或者他们可以观看具有这些演员的另一个视频内容。因此，本技术允许智能、交互式和用户特定的视频内容分发。

[0006] 根据本公开的一个示例实施例，提供了一种用于交互式视频内容分发的系统。示例系统可以驻留在基于云的计算环境中的服务器上；该系统可以与用户设备集成；或者可以直接或间接地可操作地连接到用户设备。该系统可以包括通信模块，所述通信模块被配置为接收视频内容，所述视频内容包括一个或多个视频帧。该系统还可以包括视频分析器模块，所述视频分析器模块被配置为对所述一个或多个视频帧运行一个或多个机器学习分类器，以创建分类元数据以及与所述分类元数据相关联的一个或多个概率得分，所述分类元数据对应于所述一个或多个机器学习分类器。该系统还可以包括处理模块，所述处理模块被配置为基于规则集创建一个或多个交互触发器。交互触发器可以被配置为基于分类元数据和可选地基于一个或多个概率得分来触发与视频内容有关的一个或多个动作。

[0007] 根据本发明的另一示例实施例，提供了一种用于交互式视频内容分发的方法。示

例方法包括:接收包括一个或多个视频帧的视频内容;对一个或多个视频帧运行一个或多个机器学习分类器,以创建分类元数据以及与所述分类元数据相关联的一个或多个概率得分,所述分类元数据对应于所述一个或多个机器学习分类器;基于规则集创建一个或多个交互触发器;确定用于触发至少一个触发器的条件被满足;以及基于该确定、分类元数据和概率得分来触发与视频内容有关的一个或多个动作。

[0008] 在其他实施例中,方法步骤被存储在包括计算机指令的机器可读介质上,当由计算机实现时,所述计算机指令执行该方法步骤。在又一示例实施例中,硬件系统或设备可适于执行所述的方法步骤。下面描述其他特征、示例和实施例。

附图说明

[0009] 在附图的图形中,以示例的方式而不是通过限制来说明实施例,其中,类似的参考符号表示相似的元件。

[0010] 图1示出了根据一个示例实施例的用于交互式视频内容分发的示例性系统架构。

[0011] 图2示出了根据另一示例实施例的用于交互式视频内容分发的示例性系统架构。

[0012] 图3是示出根据示例实施例的用于交互式视频内容分发的方法的处理流程图。

[0013] 图4示出了根据示例实施例的用户设备的示例图形用户界面,在该界面上可以显示视频内容(例如,电影)的帧。

[0014] 图5示出了根据一个实施例的显示附加视频内容选项的用户设备的示例图形用户界面,该附加视频内容选项包括在图4的图形用户界面中呈现的叠加信息。

[0015] 图6是以计算机系统的形式示出的示例机器的示意图,在该计算机系统中执行使机器执行本文所讨论的任何一种或多种方法的指令集。

具体实施方式

[0016] 以下的详细描述包括对构成具体实施方式的一部分的附图的参考。附图示出了根据示例实施例的图示。这些示例性实施例(在本文中也被称为“示例”)被足够详细地描述以使本领域技术人员能够实践本主题。可以组合实施例,可以使用其他实施例,或者可以在不脱离权利要求的范围的情况下进行结构、逻辑和电气方面的改变。因此,下面的详细描述不在限制意义上,并且范围由所附权利要求及其等价物限定。

[0017] 本文公开的实施例的技术可以使用多种技术来实现。例如,本文所述的方法在计算机系统上执行的软件中实现,或在利用微处理器或其它专门设计的专用集成电路(ASIC)、可编程逻辑器件或其各种组合的硬件中实现。具体而言,本文所述的方法由驻留在诸如磁盘驱动器或计算机可读介质的存储介质上的一系列计算机可执行指令来实现。应当注意,本文公开的方法可以通过蜂窝电话、智能电话、计算机(例如,台式计算机、平板计算机、膝上型计算机)、游戏机、手持游戏设备等来实现。

[0018] 本发明的技术涉及所公开的用于沉浸式交互发现体验的系统和方法。该技术可供云端(over-the-top)互联网电视(如PlayStation Vue®)、在线电影和电视节目分发服务、点播流式视频和音乐服务或任何其他分发和内容分发网络(CDN)的用户使用。此外,该技术可应用于用户生成的内容(例如,直接视频上传和屏幕录制)。

[0019] 一般而言,本技术提供:从视频内容或其部分缓冲帧、分析视频内容的帧以确定关

联分类、根据规则集评估相关分类、以及基于评估激活动作。视频内容可以包括任何形式的媒体，包括但不限于直播流、基于订阅的流服务、电影、电视、互联网视频、用户生成的视频内容(例如，直接视频上传或屏幕录制)等。该技术可以允许在向用户显示预取帧之前处理视频内容并触发动作。多个分类器(例如，图像识别模块)可用于分析每个缓冲的帧，并且动态地自动检测存在于与分类相关联的帧中的一个或多个资产。

[0020] 资产类型可包括演员、地标、特效、产品、可购买项目、对象、食物或其他可检测的资产，诸如裸体、暴力、血腥、武器、亵渎、情绪、颜色等。每个分类器可以基于一个或多个机器学习算法，机器学习算法包括卷积神经网络，并且可以生成与一个或多个资产类型相关联的分类元数据。分类元数据可以指示例如是否在视频内容中检测到某些资产、关于检测到的资产的某些信息(例如，演员的身份、导演、类型、产品、产品类别、特效的类型等)、帧中检测到的资产的坐标或边界框、或者检测到的资产的大小(例如，画面中出现的暴力或血腥的程度等)。

[0021] 控件可以包装在每个分类的周围，每个分类基于规则集(预定义或动态创建的)触发特定动作。规则集可以是帧中检测到的资产、以及视频内容的其他分类元数据、观众(观看或收听的人)、一天中的时间、环境噪声、环境参数和其他合适的输入的函数。可以根据环境因素进一步定制规则集，例如位置、用户组或媒体类型。例如，当孩子在场时，父母可能希望不要展示裸体。在该示例中，系统可以描述观看环境，确定观看所显示的视频流的用户的特性(例如，确定是否存在孩子)，在预缓冲帧中检测裸体，并且在显示之前修正(例如，暂停、编辑或模糊)帧，以使裸体不被显示。

[0022] 动作还可以包括资产模糊(例如，删除、覆盖对象、模糊等)、跳过帧、调整音量、警告用户、通知用户、请求设置、提供相关信息、生成查询和执行对相关信息或广告的搜索、打开相关的软件应用程序等。缓冲和帧分析可以在接近实时的情况下执行，或者，在非现场电影或电视节目的情况下，可以在视频内容流上载到分发网络之前提前预处理。在各种实施例中，图像识别模块可以布置在基于云计算的环境中的中央服务器上，并且可以对从客户端接收到的视频内容的帧、客户端播放的镜像视频流的帧(当视频与流并行处理时)、或者被发送到客户端的视频流的帧执行分析。

[0023] 本公开的系统和方法还可以包括跟踪用户的遍历历史，并从一个或多个入口点提供针对视频内容或特定帧的用户相关信息的图形用户界面(GUI)。呈现各种相关信息的入口点的示例可以包括暂停视频内容流、选择特定视频内容、接收用户输入、检测用户姿势、接收搜索查询、语音命令等。相关信息可以包括演员信息(例如，传记和/或职业描述)、类似的媒体内容(例如，类似的电影)、相关广告、产品、计算机游戏或基于视频内容的帧或其他元数据的分析的其他合适的信息。相关信息的每一项可以被构造为一个节点。响应于接收到用户对节点的选择，与所选节点相关的信息可以呈现给用户。系统可以跨多个用户选择的节点进行跟踪遍历，并基于遍历历史生成用户简档。系统还可以记录与触发入口点相关联的帧。用户简档还可用于确定用户偏好和动作模式，以预测用户需求并基于用户简档提供与特定用户相关的信息或动作选项。

[0024] 以下实施例的详细描述包括对构成详细描述的部分的附图的参考。注意，这里描述的实施例的特征、结构或特性可以以任何合适的方式组合在一个或多个实现中。在即时描述中，提供许多具体细节，例如编程示例、软件模块、用户选择、网络事务、硬件模块、硬件

电路、硬件芯片等，以提供对实施例的透彻理解。然而，相关领域的技术人员将认识到，实施例可以在没有一个或多个特定细节的情况下实施，或者使用其他方法、组件、材料等实施。在其它实例中，不详细示出或描述公知的结构、材料或操作，以避免混淆本发明的各个方面。

[0025] 现在将参考附图来呈现本发明的实施例，附图示出了块、组件、电路、步骤、操作、处理、算法等，为简单起见统称为“元件”。这些元件可以使用电子硬件、计算机软件或其任何组合来实现。这些元件是作为硬件还是软件实现取决于对整个系统施加的特定应用和设计约束。举例来说，元件或元件的任何部分、或者元件的任何组合可以用包括一个或多个处理器的“计算系统”来实现。处理器的实例包括微处理器、微控制器、中央处理器(CPU)、数字信号处理器(DSP)、现场可编程门阵列(FPGA)、可编程逻辑器件(PLD)、状态机、门控逻辑、离散硬件电路、以及经配置以执行本公开所述的各种功能的其他合适的硬件。处理系统中的一个或多个处理器可以执行软件、固件或中间件(统称为“软件”)。无论是指软件、固件、中间件、微码、硬件描述语言或其他，术语“软件”都应广义地解释为指处理器可执行指令、指令集、代码段、程序代码、程序、子程序、软件组件、应用程序、软件应用程序、软件包、例程、副程序、对象、可执行文件、执行线程、过程、函数等。

[0026] 因此，在一个或多个实施例中，可以在硬件、软件、或其任何组合中实现本文所述的这些功能。如果在软件中实现，则这些功能可以存储在非暂时性计算机可读介质上或在非暂时性计算机可读介质上被编码为一个或多个指令或代码。计算机可读介质包括计算机存储介质。存储介质可以是计算机可以访问的任何可用介质。作为示例而非限制，这种计算机可读介质可以包括随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程ROM(EEPROM)、光盘ROM(CD-ROM)或其它光盘存储器、磁盘存储器、固态存储器或任何其它数据存储设备、上述类型的计算机可读介质的组合、或可用于以计算机可访问的指令或数据结构的形式存储计算机可执行代码的任何其他介质的组合。

[0027] 出于本专利文件的目的，术语“或”和“和”应指“和/或”，除非另有说明或在使用上下文中另有明确意图。术语“一”应指“一个或多个”，除非另有说明或使用“一个或多个”显然不合适。术语“包含”、“由…组成”、“包括”和“包括…在内”是可互换的，并不意在限制。例如，术语“包括”应解释为“包括但不限于”。术语“或”用于指代非排他性的“或”，使得“A或B”包括“A而不是B”、“B而不是A”和“A和B”，除非另有说明。

[0028] 术语“视频内容”可以指可以显示、播放和/或流式传输到下面定义的用户设备的任何类型的视听媒体。视频内容的一些示例包括但不限于视频流、直播流、电视节目、直播电视、点播视频、电影、影片、动画、互联网视频、多媒体、视频游戏、计算机游戏等。视频内容可以包括用户生成的内容，例如，直接视频上传和屏幕录制。术语“视频内容”、“视频流”、“媒体内容”和“多媒体内容”可以互换使用。视频内容包括多个帧(视频帧)。

[0029] 术语“用户设备”可以指能够接收和向用户呈现视频内容的设备。用户设备的一些示例包括但不限于电视设备、智能电视系统、计算设备(例如，平板电脑、笔记本电脑、台式计算机或智能电话)、投影电视系统、数字视频录像机(DVR)设备、游戏机、游戏设备、多媒体系统娱乐系统、计算机实现的视频回放设备、移动多媒体设备、移动游戏设备、机顶盒(STB)设备、虚拟现实设备、数字视频录像机(DVR)、远程存储DVR等。STB设备可以部署在用户的家庭中，以向用户提供对从内容提供者分发的视频内容的交互控制的能力。术语“用户”、“观

看者”、“观众”和“玩家”可以互换地用于表示使用如上所定义的用户设备的人，或者表示如本文所述观看视频内容的人。用户可以通过提供用户输入或用户姿势与用户设备进行交互。

[0030] 术语“分类元数据”是指与一个或多个资产或诸如视频内容对象或特性的电子内容项相关联(并且通常，但不一定与之一起存储)的信息。术语“资产”是指视频内容的项目，例如包括视频内容中包含的或与视频内容相关联的对象、文本、图像、视频、音频、个体、参数或特性。分类元数据可以包含唯一标识资产的信息。这种分类元数据可以描述资产的存储位置或其他唯一标识。例如，与出现在视频内容的某些帧中的演员相关联的分类元数据可以包括名称和/或标识符，或者可以以其他方式描述与演员相关的附加内容(或链接)的存储位置。

[0031] 现在参考附图描述示例实施例。附图是理想化示例实施例的示意图。因此，本文讨论的示例性实施例不应被解释为限于本文所呈现的特定图示。并且可以包括与本文中的示例不同的示例。

[0032] 图1示出了根据一个示例实施例的用于交互式视频内容分发的示例性系统架构100。系统架构100包括交互式视频内容分发系统105、一个或多个用户设备110和一个或多个内容提供者115。例如，可以通过一个或多个计算机服务器或基于云的服务来实现系统105。用户设备110可以包括电视设备、STB、计算设备、游戏机等。同样地，用户设备110可以包括输入和输出模块，以使用户能够控制视频内容的回放。视频内容可以由诸如内容服务器、视频流服务、互联网视频服务或电视广播服务的一个或多个内容提供者115提供。视频内容可以由用户生成，例如，作为直接视频上传或屏幕录制。术语“内容提供者”可以被广义地解释为包括可以参与使用户能够经由用户设备110获得对特定内容的访问的处理的任何当事人、实体、设备或系统。内容提供者115还可以表示或包括内容分发网络(CDN)。

[0033] 交互式视频内容分发系统105、用户设备110和内容提供者115可以经由通信网络120可操作地彼此连接。通信网络120可以指任何有线、无线或光网络，包括例如因特网、内联网、局域网(LAN)、个人局域网(PAN)、广域网(WAN)、虚拟专用网(VPN)、蜂窝电话网络(例如，分组交换通信网，电路交换通信网络)、蓝牙无线电、以太网网络、基于IEEE 802.11的射频网络、IP通信网络或利用物理层、链路层能力或网络层来承载数据包的任何其他数据通信网络、或上述数据网络的任何组合。

[0034] 交互式视频内容分发系统105可以包括至少一个处理器和用于存储与本文公开的方法相关联的处理器可执行指令的至少一个存储器。如图所示，交互式视频内容分发系统105包括各种模块，这些模块可以在硬件、软件或两者中实现。同样地，交互式视频内容分发系统105包括用于从内容提供者115接收视频内容的通信模块125。通信模块125还可以向用户设备110或内容提供者115发送视频内容、编辑的视频内容、分类元数据或与用户或视频内容相关联的其他数据。

[0035] 交互式视频内容分发系统105还可以包括视频分析器模块130，所述视频分析器模块130被配置为对经由通信模块125接收到的视频内容的视频帧运行一个或多个机器学习分类器。机器学习分类器可以包括神经网络、深度学习系统、启发式系统、统计数据系统等。如下文所述，机器学习分类器可包括一般对象分类器、产品分类器、环境条件分类器、情绪条件分类器、地标分类器、人物分类器、食物分类器、问题内容分类器等。视频分析器模块

130可以并行地并且彼此独立地运行上述机器学习分类器。

[0036] 上述分类器可以包括图像识别分类器或复合识别分类器。图像识别分类器可以被配置为分析一个或多个视频帧中的静态图像。复合识别分类器可以被配置为分析:(i)两个或多个视频帧之间的一个或多个图像变化;以及(ii)两个或多个视频帧之间的一个或多个声音变化。作为输出,上述分类器可以创建对应于一个或多个机器学习分类器的分类元数据以及与该分类元数据相关联的一个或多个概率得分。概率得分可以参考特定视频帧包括或与特定资产(例如,出现在视频帧中的演员、对象或可购买项目)相关联的置信水平(例如,因子、权重)。

[0037] 在一些实施例中,视频分析器模块130可以通过将内容分发缓冲并延迟处理实时视频的视频帧所需的时间来执行实时视频内容的分析。在其他实施例中,视频分析器模块130可以执行用于按需分发的视频内容的分析。如上所述,实时视频内容可以被缓冲在交互式视频内容分发系统105的存储器中,使得视频内容以微小的延迟被分发并呈现给用户,以使得视频分析器模块130能够执行视频内容的分类。

[0038] 交互式视频内容分发系统105还可以包括处理模块135,所述处理模块135被配置为基于规则集来创建一个或多个交互触发器。交互触发器可以被配置为基于分类元数据和(可选地)概率得分来触发关于视频内容的一个或多个动作。可以基于以下一个或多个预定义或动态选择该规则:用户简档、用户设置、用户偏好、观看者身份、观看者年龄和环境条件。这些动作可以包括编辑视频内容(例如,编辑、模糊、突出显示、调整颜色或音频特性等)、控制视频内容的分发(例如,暂停、跳过和停止),以及呈现与视频内容相关联的附加信息(例如,警示用户、通知用户,提供有关视频内容中存在的对象、地标、人物等的附加信息,提供超链接、以及允许用户进行购买)。

[0039] 图2示出了根据另一示例实施例的用于交互式视频内容分发的示例性系统架构200。与图1类似,系统架构200包括交互式视频内容分发系统105、一个或多个用户设备110和一个或多个内容提供者115。然而,在图2中,交互式视频内容分发系统105是一个或多个用户设备110的一部分,或与一个或多个用户设备110集成。换言之,交互式视频内容分发系统105可以在用户位置处提供本地视频处理(如本文所述)。例如,交互式视频内容分发系统105可以是STB或游戏机的功能。交互式视频内容分发系统105和系统架构200的其它元件的操作和功能与上文参照图1所述的相同或基本相同。

[0040] 图2还示出与用户设备110通信地耦合的一个或多个传感器205。传感器205可以被配置为检测、确定、识别或测量与一个或多个用户、用户的家(场所)、用户的环境或周边参数等相关联的各种参数。传感器205的一些示例包括摄像机、麦克风、运动传感器、深度照相机、光电探测器等。例如,传感器205可用于检测和识别用户、确定儿童是否观看或访问特定视频内容、确定照明条件、测量噪声水平、跟踪用户行为、检测用户情绪等。

[0041] 图3是示出根据示例实施例的用于交互式视频内容分发的方法300的处理流程图。方法300可以通过包括硬件(例如,决策逻辑、专用逻辑、可编程逻辑、专用集成电路)、软件(例如在通用计算机系统或专用机器上运行的软件)或二者的组合的处理逻辑来实现。在示例性实施例中,处理逻辑涉及图1和2的交互式视频内容分发系统105的一个或多个元件。下面所述的方法300的操作可以不同于图中描述和示出的顺序来实现。此外,方法300可以具有本文未示出的附加操作,但是对于本领域技术人员来说,可以从本公开中明显看出。方法

300也可以具有比图3所示和下面描述的更少的操作。

[0042] 方法300从操作305开始，其中通信模块125接收视频内容，视频内容包括一个或多个视频帧。可以从一个或多个内容提供者115、CDN或本地数据存储器接收视频内容。如上所述，视频内容可以包括多媒体内容(例如，电影、电视节目、点播视频、音频、点播音频)、游戏内容、体育内容、音频内容等。视频内容可以包括实时流或预录制的内容。

[0043] 在操作310处，处理模块130可以对一个或多个视频帧运行一个或多个机器学习分类器，以创建与一个或多个机器学习分类器相对应的分类元数据，以及与分类元数据相关联的一个或多个概率得分。机器学习分类器可以并行运行。另外，可以在将视频内容上载到CDN、内容提供者115或流式传输到用户或用户设备110之前在视频内容上运行机器学习分类器。

[0044] 分类元数据可以表示视频内容的一个或多个资产、周边或环境条件、用户信息等，或与视频内容的一个或多个资产、周边或环境条件、用户信息等相关联。视频内容的资产可以与对象、人物(例如，演员、电影导演等)、食物、地标、音乐、音频项目或视频内容中存在的其他项目相关。

[0045] 在操作315处，处理模块135可以基于规则集创建一个或多个交互触发器。交互触发器被配置为基于分类元数据并且可选地基于一个或多个概率得分来触发关于视频内容的一个或多个动作。规则集可以基于以下一个或多个：用户简档、用户设置、用户偏好、观看者身份、观看者年龄和环境条件。在一些实施例中，可以预定义规则集。在其他实施例中，可以动态地创建、更新或选择规则集以反映用户偏好、用户行为或其他相关情况。

[0046] 在操作320处，用户设备110向一个或多个用户呈现视频内容。在执行操作305-315之后，可以对视频内容进行流式处理。在操作320呈现视频内容时，用户设备110可以通过传感器205测量一个或多个参数。

[0047] 在操作325处，交互式视频内容系统105或用户设备110可以确定用于触发至少一个或多个交互触发器的条件被满足。所述条件可以是预定义的，并且可以是多个条件中的一个。在一些实施例中，条件是指入口点或与入口点相关联。在方法300中，交互式视频内容系统105或系统架构100或200的任何其他元件可以创建对应于交互触发器的一个或多个入口点。每个入口点包括与视频内容相关联的用户输入，或与视频内容相关联的用户姿势。具体地说，每个入口点可以包括以下一个或多个：视频内容的暂停、视频内容的跳转点、视频内容的书签、视频内容的位置标记、由所连接的传感器检测到的用户环境的变化、以及与视频内容相关联的搜索结果。换句话说，在一个示例实施例中，操作325可以确定用户是否暂停了视频内容、按下了预定按钮，或者内容是否到达了位置标记。在另一个示例实施例中，操作325可以利用用户设备110上的传感器来确定用户环境的变化是否创造了触发交互触发器的条件。例如，用户设备110上的照相机传感器可以确定儿童何时走进房间，并且交互式视频内容系统105或用户设备110可以自动模糊问题内容(例如，可能不适合儿童的内容)。此外，另一传感器驱动的入口点可以包括语音控制(即，用户可以使用连接到用户设备110的麦克风来询问“屏幕上的演员是谁?”作为响应，交互式视频内容系统105或用户设备110可以响应于用户的询问来呈现数据。

[0048] 在操作330处，交互视频内容系统105或用户设备110响应于在操作325所做的确定触发关于视频内容的一个或多个动作。在一些实施例中，动作可以基于与视频内容的入口

点中的一个相关联的帧的分类元数据。通常，动作可以涉及提供附加信息、视频内容选项、链接(超链接)、突出显示、修改视频内容、控制视频内容的回放等。动作可以取决于分类元数据(即，基于生成元数据的机器学习分类器)。应该理解，交互触发器可以在主屏幕或副屏幕上显示信息和动作。例如，地标的名称可以显示在与主屏幕上的帧相匹配的设备(如智能手机)上。在另一示例中，副屏幕可以在主屏幕上显示正在观看的帧中的可购买项目，从而允许在副屏幕上直接购买项目。

[0049] 在各种实施例中，机器学习分类器中的每一个可以是至少两种类型：(i)图像识别分类器，其被配置为分析视频帧的一个中的静态图像，以及(ii)符合识别分类器，其被配置为分析：(a)两个或多个视频帧之间的一个或多个图像变化；以及(b)两个或多个视频帧之间的一个或多个声音变化。

[0050] 一个实施例提供一般对象分类器，其被配置为识别在一个或多个视频帧中存在的一个或多个对象。对于该分类器，在触发一个或多个交互触发器时要采取的动作可以包括以下一个或多个：用视频帧中的新对象替换对象、自动突出显示对象、推荐对象表示的可购买项目、基于对象的标识编辑视频内容、基于对象的标识控制视频内容的分发、以及呈现与对象相关的搜索选项。

[0051] 另一实施例提供产品分类器，其被配置为识别视频帧中存在的一个或多个可购买项目。对于该分类器，在触发一个或多个交互触发器时要采取的动作可以包括，例如，提供一个或多个链接以使用户能够购买一个或多个可购买项目。

[0052] 又一实施例提供环境条件分类器，其被配置为确定与视频帧相关联的环境条件。这里，可以基于以下传感器数据来创建分类元数据：一个或多个观看者正在观看视频内容的场所的照明条件、场所的噪声水平、与场所相关联的观众观看者类型、观看者身份和当前时间。使用一个或多个传感器205获得传感器数据。对于该分类器，在触发一个或多个交互触发器时要采取的动作包括以下一个或多个：基于环境条件编辑视频内容、基于环境条件控制视频内容的分发、基于环境条件提供与视频内容或另一媒体内容相关联的建议、以及提供与环境条件相关联的另一媒体内容。

[0053] 另一实施例提供情绪条件分类器，其被配置为确定与一个或多个视频帧相关联的情绪水平。在本实施例中，可以基于以下一个或多个来创建分类元数据：一个或多个视频帧的颜色数据、一个或多个视频帧的音频信息、以及用户响应于观看视频内容的行为。此外，在本实施例中，在触发一个或多个交互触发器时要采取的动作可以包括以下一个或多个：提供关于与情绪水平相关联的另一媒体内容的建议、以及提供与情绪水平相关联的其他媒体内容。

[0054] 一个实施例提供地标分类器，其被配置为识别在一个或多个视频帧中存在的地标。对于该分类器，在触发一个或多个交互触发器时要采取的动作可以包括以下一个或多个：在一个或多个视频帧中标记所识别的地标，提供关于与所识别的地标相关联的另一媒体内容的建议、提供与所识别的地标相关联的其他媒体内容、基于所识别的地标编辑视频内容、基于所识别的地标控制视频内容的分发、以及呈现与所识别的地标相关的搜索选项。

[0055] 另一实施例提供人物分类器，其被配置为识别视频帧中存在的一个或多个个体。对于该分类器，在触发一个或多个交互触发器时要采取的动作包括以下一个或多个：在一个或多个视频帧中标记一个或多个个体、提供关于与一个或多个个体相关联的另一媒体内

容的建议、提供与一个或多个个体相关联的其他媒体内容、基于一个或多个个体编辑视频内容、基于一个或多个个体控制视频内容的分发、以及呈现与一个或多个个体相关的搜索选项。

[0056] 又一实施例提供食物分类器，其被配置为识别在一个或多个视频帧中存在的一个或多个食物项。对于该分类器，在触发一个或多个交互触发器时要采取的动作包括以下一个或多个：在一个或多个视频帧中标记一个或多个食物项、提供与一个或多个食物项相关的营养信息、为用户提供购买与一个或多个食物项相关联的可购买项目的购买选项、提供与一个或多个食物项相关的媒体内容、以及提供与一个或多个食物项相关的搜索选项。

[0057] 一个实施例提供问题内容分类器，其被配置为检测一个或多个视频帧中的问题内容。问题内容可能包括以下一个或多个：裸体、武器、酒精、烟草、毒品、血液、仇恨言论、亵渎、血腥、以及暴力。对于该分类器，在触发一个或多个交互触发器时要采取的动作可以包括以下一个或多个：在向用户显示之前自动模糊一个或多个视频帧中的问题内容、跳过与问题内容相关联的视频内容的部分、基于问题内容编辑视频内容、基于问题内容调整视频内容的音频、基于问题内容调整音频音量水平、基于问题内容控制视频内容的分发、以及将问题内容通知用户。

[0058] 图4示出了根据一个实施例的用于显示视频内容(例如，电影)的至少一帧的用户设备110的示例图形用户界面(GUI)400。该示例GUI示出当用户暂停视频内容的回放时，由交互式视频内容系统105检测入口点。响应于该检测，交互式视频内容系统105触发与视频帧中识别的演员相关联的动作。该动作可以包括提供关于演员的叠加信息405(在本例中，示出了演员的姓名和脸部帧)。值得注意的是，关于演员的信息405可以实时动态地生成，但这不是必需的。可以基于缓冲的视频内容生成信息405。

[0059] 在一些实施例中，叠加(或覆盖)信息405可以包括超链接。叠加信息也可以用一个可动作的“软”按钮来表示。通过这样的按钮，用户可以通过用户输入或用户姿势来选择、按下、点击或以其他方式激活叠加信息405。

[0060] 图5示出了根据一个实施例的用户设备110的示例性图形用户界面500，其示出了与图4的图形用户界面400中存在的叠加信息405相关联的附加视频内容选项505。换句话说，当用户激活GUI 400中的叠加信息405时，显示GUI 500。

[0061] 如图5所示，GUI 500包括多个视频内容选项505，诸如具有与图4中识别的相同演员的电影。GUI 500还可以包括提供关于图4中识别的演员的数据的信息容器(container)510。信息容器510可以包括文本、图像、视频、多媒体、超链接等。用户还可以选择一个或多个视频内容选项505，并且这些选择可以被保存到用户简档中，以便用户可以在以后的时间访问这些视频内容选项505。另外，机器学习分类器可以监控由用户的选择表示的用户的行为，以确定用户的偏好。系统105可以进一步利用用户偏好来选择和向用户提供建议。

[0062] 图6以计算机系统600的示例电子形式示出了机器的计算设备的示意表示，在该计算机系统中执行使机器执行本文所讨论的任何一种或多种方法的指令集。在示例实施例中，该机器作为独立设备运行，或者可以连接(例如，联网)到其他机器。在网络部署中，机器可以在服务器-客户端网络环境中作为服务器或客户端机器运行，或者在对等(或分布式)网络环境中作为对等计算机运行。该机器可以是个人计算机(PC)、平板电脑、游戏机、游戏设备、机顶盒(STB)、电视设备、蜂窝电话、便携式音乐播放器(例如，便携式硬盘驱动器音频

设备)、web设备、或能够执行指定了该机器将要采取的动作的指令集(顺序或其他方式)的任何机器。此外,虽然仅示出了一台机器,但术语“机器”还应被视为包括单独或联合执行一组(或多组)指令以执行本文所讨论的任何一种或多种方法的任何机器的集合。计算机系统600可以是交互式视频内容分发系统105、用户设备110或内容提供者115的实例。

[0063] 示例性计算机系统600包括一个或多个处理器605(例如,中央处理单元(CPU)、图形处理单元(GPU),或两者)以及通过总线620彼此通信的主存储器610和静态存储器615。计算机系统600还可以包括视频显示单元625(例如,LCD)。计算机系统600还包括至少一个输入设备630,诸如字母数字输入设备(例如键盘)、光标控制设备(例如鼠标)、麦克风、数码相机、摄像机等。计算机系统600还包括磁盘驱动器单元635、信号生成设备640(例如扬声器)和网络接口设备645。

[0064] 驱动单元635(也称为磁盘驱动器单元635)包括机器可读介质650(也称为计算机可读介质650),其存储由本文描述的方法或功能中的任何一个或多个实现或使用的一组或多组指令和数据结构(例如,指令655)。在计算机系统600执行指令655期间,指令655还可以完全或至少部分地驻留在主存储器610和/或处理器605内。主存储器610和处理器605也构成机器可读介质。

[0065] 还可以利用许多已知传输协议(例如,超文本传输协议(HTTP)、CAN、串口和(网络通讯协议)Modbus)中的任何一个经由网络接口设备645在通信网络660上发送或接收指令655。通信网络660包括因特网、局域网、个人局域网(PAN)、局域网(LAN)、广域网(WAN)、城域网(MAN)、虚拟专用网(VPN)、存储区域网(SAN)、帧中继连接、高级智能网(AIN)连接、同步光网络(SONET)连接、数字T1、T3、E1或E3线路、数字数据业务(DDS)连接、数字用户线(DSL)连接、以太网连接、综合业务数字网(ISDN)线路、电缆调制解调器、异步传输模式(ATM)连接、或光纤分布式数据接口(FDDI)、或铜缆分布式数据接口(CDDI)连接。此外,通信网络660还可以包括到各种无线网络中的任何一个的链接,各种无线网络包括无线应用协议(WAP)、通用分组无线业务(GPRS)、全球移动通信系统(GSM)、码分多址(CDMA)或时分多址(TDMA)、蜂窝电话网络,全球定位系统(GPS)、蜂窝数字分组数据(CDPD)、动态研究、有限(RIM)双工寻呼网络、蓝牙无线电、或基于IEEE 802.11的射频网络。

[0066] 虽然机器可读介质650在示例实施例中被示出为单个介质,但术语“计算机可读介质”应被视为包括存储一组或多组指令的单个介质或多个介质(例如,集中式或分布式数据库,和/或相关联的缓存和服务器)。术语“计算机可读介质”还应被视为包括能够存储、编码或携带指令集以供机器执行、并使机器执行本申请的任何一种或多种方法的任何介质,或能够存储、编码或者携带由这样一组指令使用或与之相关联的数据结构的任何介质。术语“计算机可读介质”应相应地包括但不限于固态存储器、光学和磁性介质。这种介质还可以包括但不限于硬盘、软盘、闪存卡、数字视频盘、随机存取存储器(RAM)、只读存储器(ROM)等。

[0067] 本文描述的示例性实施例可以在包括计算机可执行指令(例如,软件)的操作环境中实现,所述可执行指令安装在计算机上、硬件中或软硬件组合中。计算机可执行指令可以用计算机编程语言编写,也可以体现在固件逻辑中。如果用符合公认标准的编程语言编写,则这些指令可以在各种硬件平台上执行,并且也可以用于与各种操作系统的接口。尽管不限于此,用于实现本方法的计算机软件程序可以用任何数量的合适的编程语言编写,例如,

超文本标记语言(HTML)、动态HTML、XML、可扩展样式表语言(XSL)、文档样式语义和规范语言(DSSSL)，级联样式表(CSS)、同步多媒体集成语言(SMIL)、无线标记语言(WML)、Java™、Jini™、C、C++、C#、.NET、Adobe Flash、Perl、UNIX Shell、Visual Basic或Visual Basic脚本、虚拟现实标记语言(VRML)、ColdFusion™或其他编译器、汇编器、解释器，或其他计算机语言或平台。

[0068] 因此，公开了用于交互式视频内容分发的技术。尽管已经参考具体示例实施例描述了实施例，但是显而易见的是，可以在不脱离本申请的更广泛的精神和范围的情况下对这些示例实施例进行各种修改和更改。因此，说明书和附图应被视为说明性而非限制性意义。

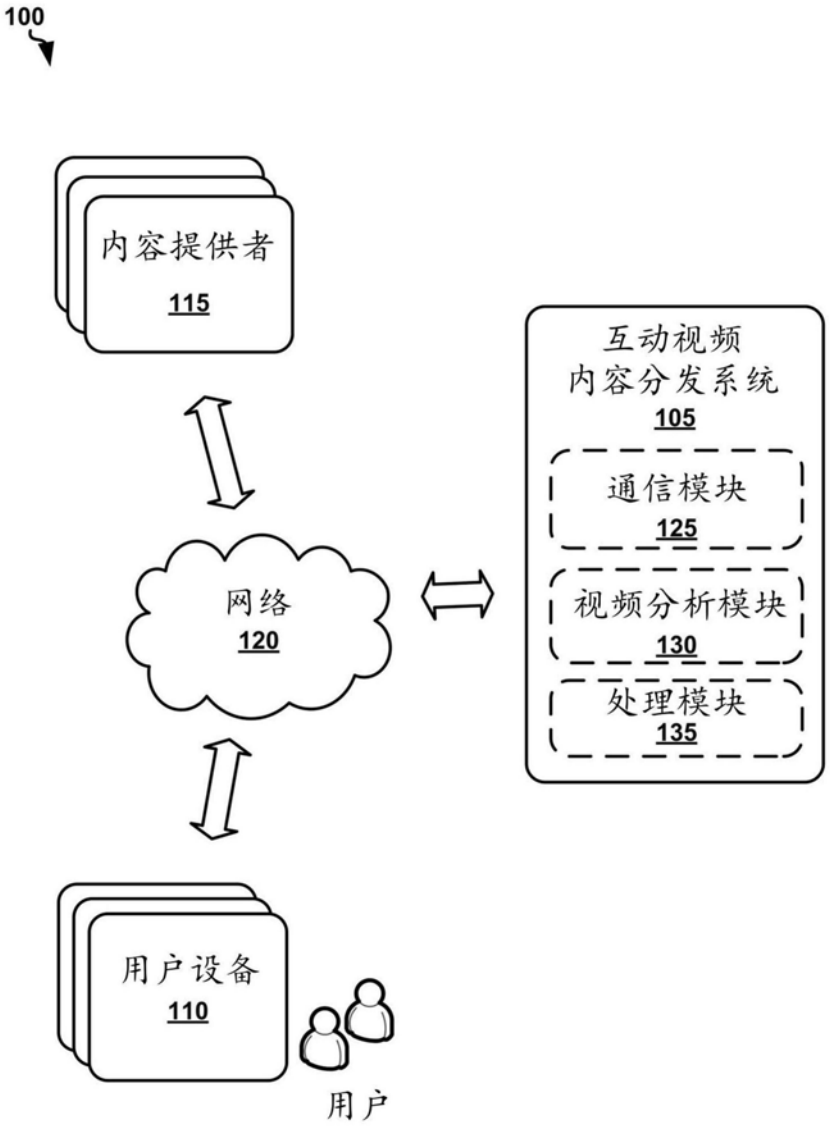

100
内容提供者
115
互动视频
内容分发系统
105
通信模块
125
网络
120
视频分析模块
130
处理模块
135
用户设备
110
用户

图1

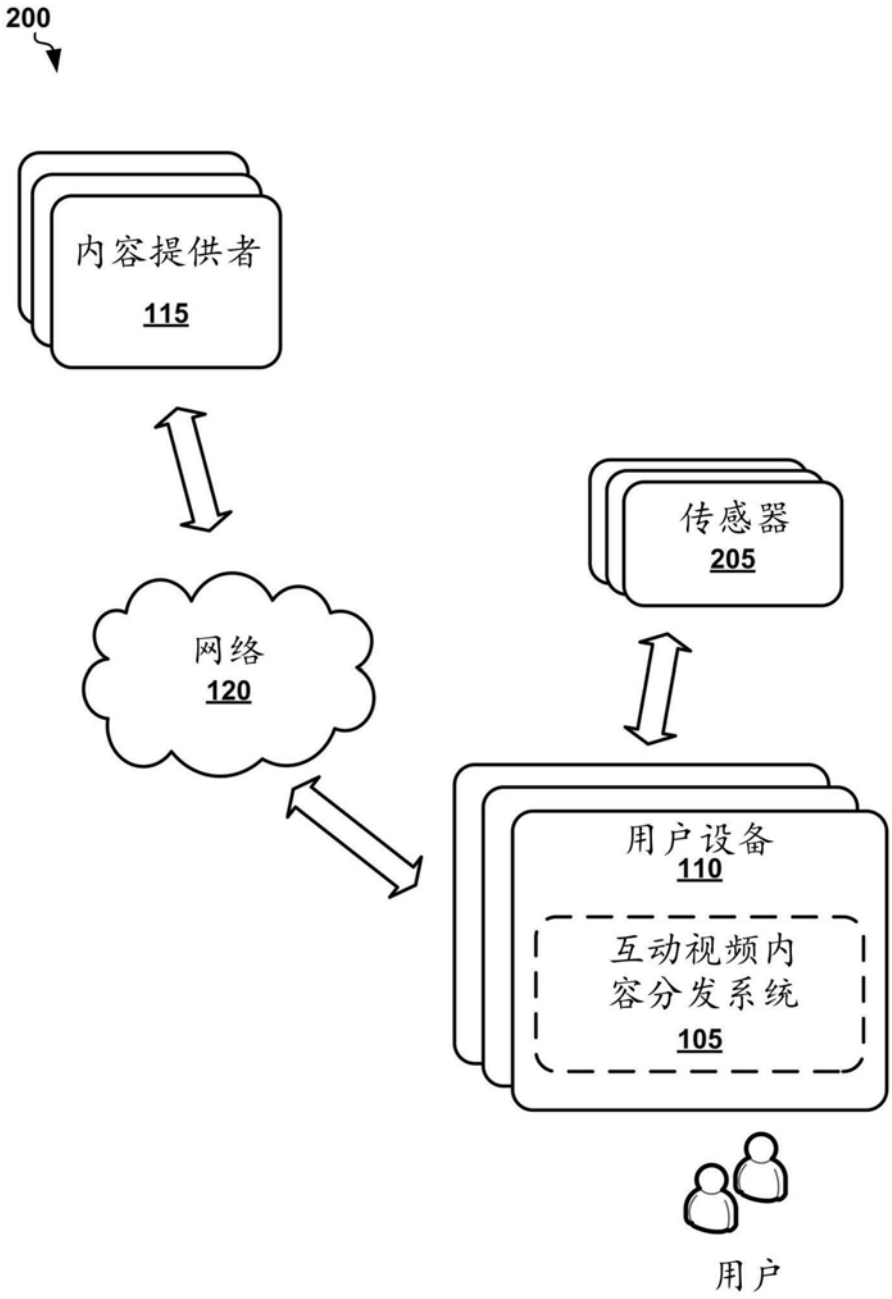


图2

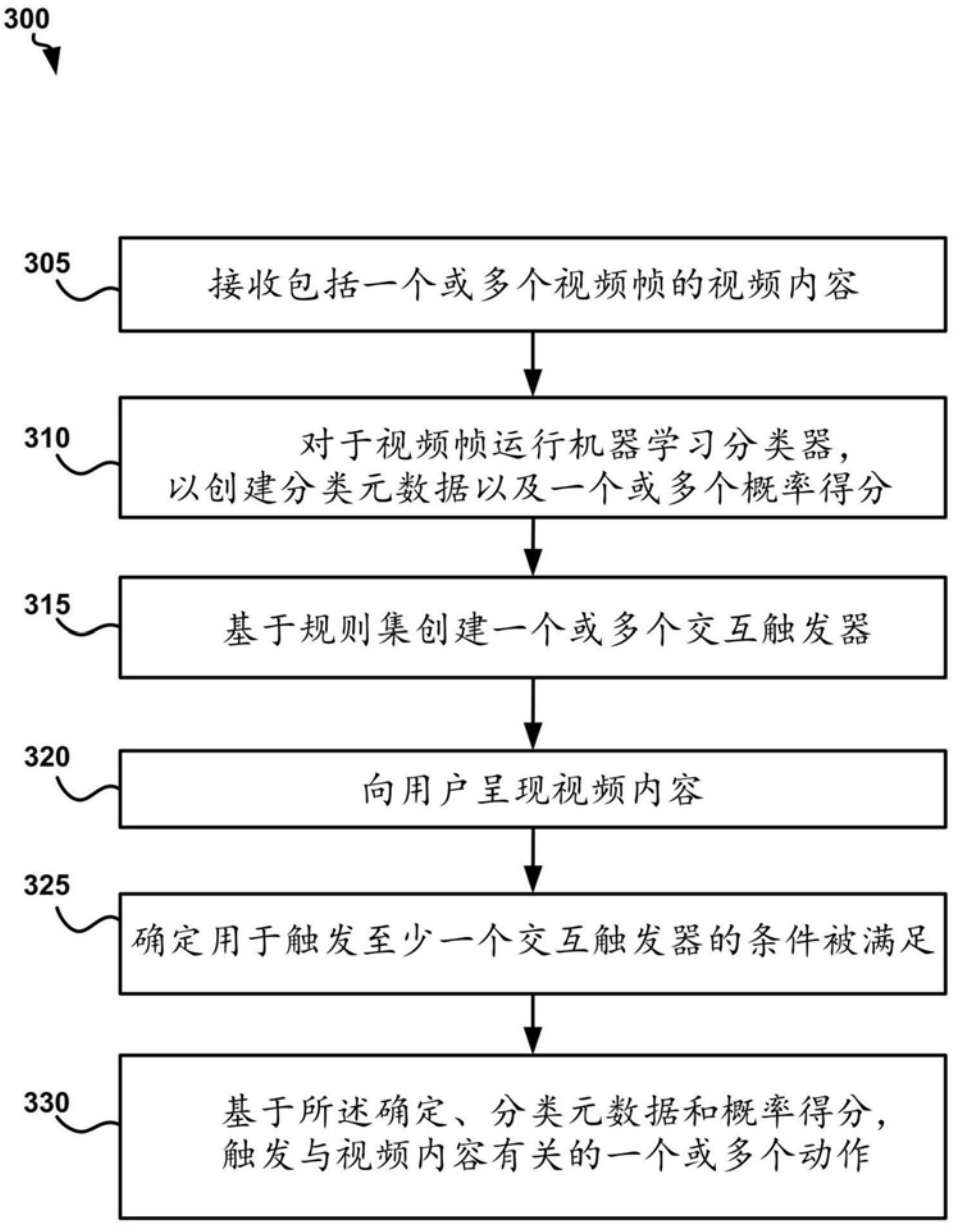

300
305
接收包括一个或多个视频帧的视频内容
310
对于视频帧运行机器学习分类器，
以创建分类元数据以及一个或多个概率得分
315
基于规则集创建一个或多个交互触发器
320
向用户呈现视频内容
325
确定用于触发至少一个交互触发器的条件被满足
330
基于所述确定、分类元数据和概率得分，
触发与视频内容有关的一个或多个动作

图3

图4

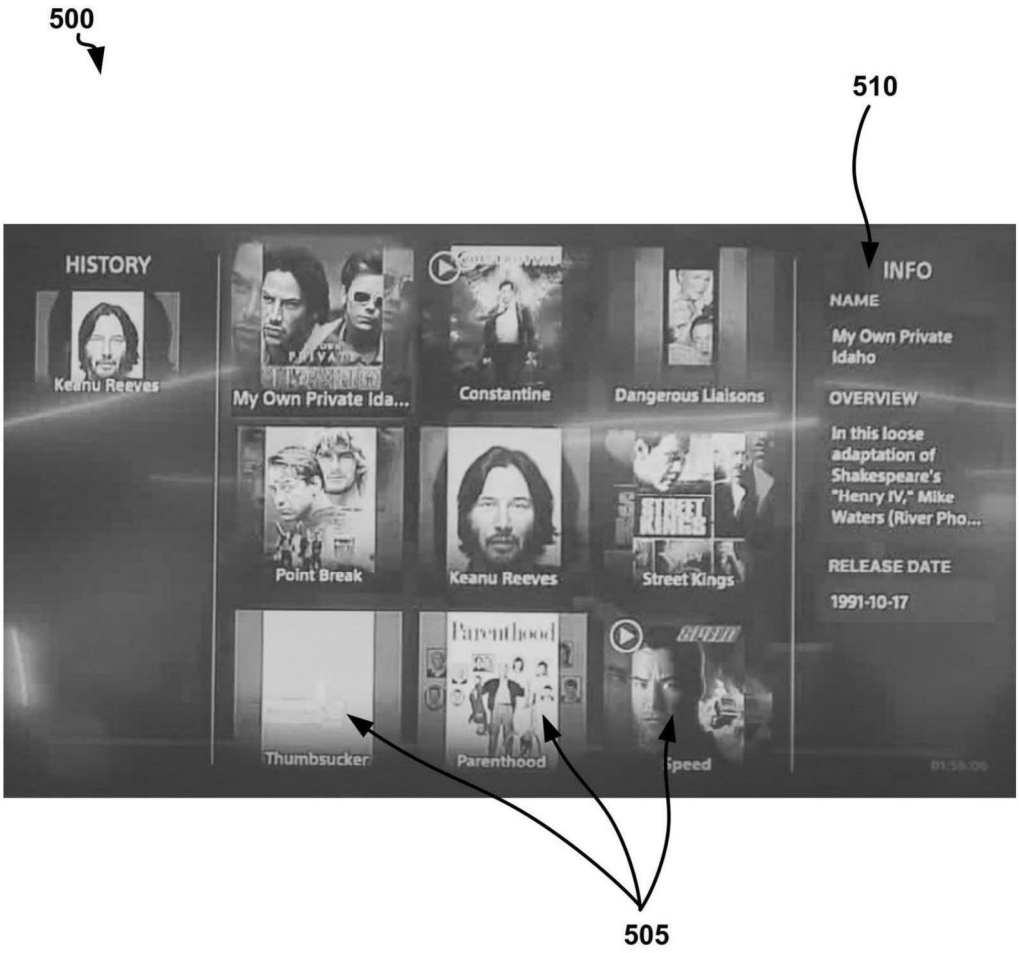

500
510
HISTORY
Keanu Reeves
My Own Private Ida...
Constantine
Dangerous Liaisons
INFO
NAME
My Own Private Idaho
OVERVIEW
In this loose adaptation of Shakespeare's "Henry IV," Mike Waters (River Pho...
RELEASE DATE
1991-10-17
Point Break
Keanu Reeves
Street Kings
Thumbsucker
Parenthood
Speed
505

图5

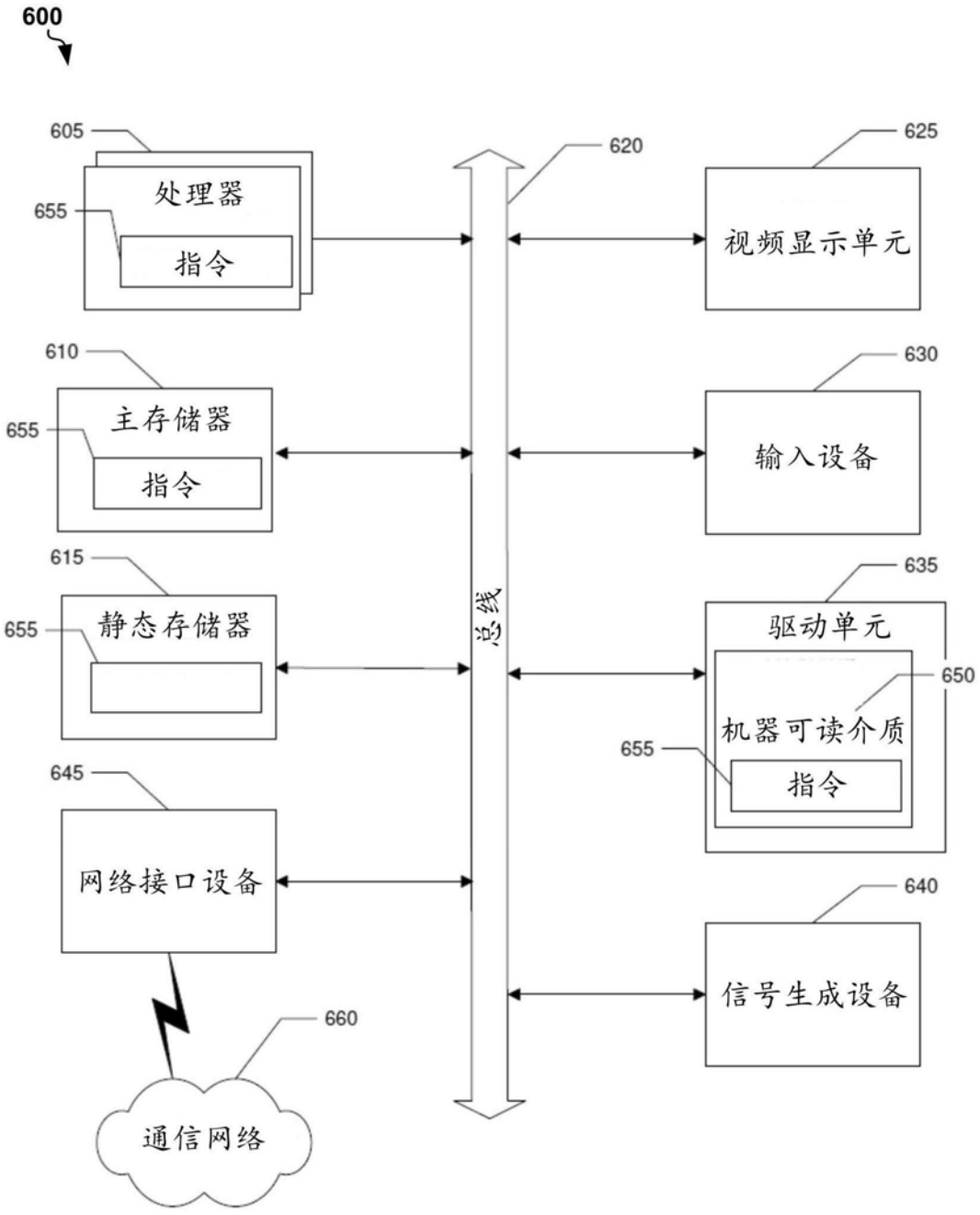

600
605
处理器
655
指令
620
625
视频显示单元
610
主存储器
655
指令
630
输入设备
615
静态存储器
655
总线
635
驱动单元
650
机器可读介质
655
指令
645
网络接口设备
640
信号生成设备
660
通信网络

图6